DOCUMENT RESUME

ED 423 259 TM 028 994

AUTHOR Swearingen, Dorothy L.
TITLE Person Fit and Its Relationship with Other Measures of Response Set.
PUB DATE 1998-04-00
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Attitude Measures; *College Students; Higher Education; Item Response Theory; *Measurement Techniques; *Response Style (Tests); Semantics; Test Construction; *Test Items
IDENTIFIERS BIGSTEPS Computer Program; *Person Fit Measures; *Rasch Model

ABSTRACT

When response set is present, instead of responding to the intent of the question, the subject appears to be responding to a variable emanating from some personal characteristic. This threat to measurement reliability and validity warrants investigation of the source of response set so that questionnaire designers can minimize its occurrence. This study sought to identify response sets most closely associated with person fit, which has been shown to be an effective method for identifying response sets on a questionnaire. Subjects were 597 undergraduate and graduate students who were administered a thinking style measure and an attitude questionnaire on 2 controversial topics, abortion and homosexual rights, and 2 noncontroversial questions, arts education and standardized questions. Three item formats were used. The BIGSTEPS computer program was used to measure individual misfit, and when person fit and other response sets were found in the correlational analysis to be highly associated, verification was sought in the Rasch output. The moderate-to-substantial correlations between infit and extreme responding style and between infit and response range found on the semantic differential (SD), and rating scale (RS) item formats were not seen for the magnitude estimation scale (ME), suggesting that fit statistics may be useful in determining response set on the SD and RD scales for all but the acquiescence/directional (AD) set, but perhaps is not as useful for the ME scale. Because of the high associations observed, the measurement of person fit through use of the Rasch model is an effective method for determining response set. (Contains 4 tables, 9 figures, and 28 references.) (SLD)

PERSON FIT AND ITS RELATIONSHIP WITH OTHER MEASURES OF RESPONSE SET

Dorothy L. Swearingen, Ph.D.
1440 South Garfield Street
Denver, CO 80210-2534
email: dswearin@ix.netcom.com

Paper presented at the annual meeting of the American Educational Research Association (AERA), San Diego, CA, April, 1998

Introduction

The problem of response set has plagued interpreters of questionnaires for decades. As early as 1925, Allport and Hartmann (cited in Cantril, 1946) were attempting to identify sources of this phenomenon. Measurement characteristics — such as questionnaire length, item format, item content, use of a midpoint, number of response categories -- and personal characteristics -- such as ethnicity, gender, certainty, thinking style, personality -- have been investigated to help identify variables responsible for this threat to reliability and validity in measurement (Alwin & Krosnick, 1991; Bachman & O'Malley, 1984; Cronbach, 1946, 1950; Edwards, 1953; Hamilton, 1968; Hui & Triandis, 1985, 1989; Rorer, 1965; Swearingen, 1997).

Definitions of response set are varied. Cronbach (1946) defined it as a response to items that is consistently different from the person's response to the same items in another form. He found it most problematic with instruments measuring personality, attitude, interest, and ability. Edwards (1953) believed it to be related to a personal need to create a specific impression. Hui and Triandis (1985) define it as a "tendency to respond in a manner that is unrelated to the content of the instrument" (p. 253). Hamilton (1968) portrays it as consistent and uniquely personal. Though opinions vary as to its definition, the elements of consistency and independence from the content of the items on a questionnaire have been generally accepted. Swearingen (1997), however, in a study examining the effects of item format, item controversy, and thinking style on response set, found controversy of content to be a significant contributor.

When response set is present, instead of responding to the intent of the questions, the subject appears to be responding to a variable emanating from some personal characteristic. This threat to measurement reliability and validity warrants ongoing investigation of sources of response set so that questionnaire designers can minimize its occurrence.

Response set is most directly a problem for interpreters of questionnaires, who may draw the wrong conclusions from their research, or who may find they have to drop significant numbers of subjects from their data due to responses they consider invalid. However, response set becomes a problem for the public as well when unsupportable conclusions are derived from research. For example, leaders in education, business, and government often make policy decisions based on surveys. Decisions having a basis in error can lead to a decline in production or profits, or a loss of support from essential participants.

Several models have been developed to help us identify response set. The most widely researched sets are: 1) the social desirability response set (Beardon & Rose, 1990; Edwards, 1953; Meisels & Ford, 1969); and 2) the extreme responding style (Allport & Hartmann, 1925, cited in Cantril, 1946; Bachman & O'Malley, 1984; Hui

& Triandis, 1985, 1989; White & Harvey, 1965). Other patterns that have been identified are: 1) acquiescence/directional bias (Cronbach, 1946, 1950; Hui & Triandis, 1985; McClendon, 1991; Rorer, 1965); 2) response range (Hui & Triandis, 1985; Wilcox, Sigelman, & Cook, 1989); 3) primacy and recency effects (Tittle & Hill, 1967); and 4) scatter and ratings (Schnellbecker, 1993). However, in addition to the conventional response sets, a statistic called person fit, derived from analysis using the Rasch model, may offer additional information on several response sets.

Person fit refers to the believability of a person's pattern of response on an assessment measure (Smith, 1986), given the person's ability (independent of items) and the item's difficulty (independent of persons). Both person ability and item difficulty are placed on a common scale, expressed in logits, with an expected mean value of 1.0 and a standard deviation of 0. A person's ability represents his/her log odds for succeeding on an item with difficulty of zero, or mean difficulty (Wright & Stone, 1979). By examining the difference between ability and difficulty, an estimation of a person's expected response to an item can be made. When expected and observed responses are compared, using the Rasch method, person fit statistics, expressed as standardized mean squares, are derived.

With an attitude measure, the focus is not on a level of ability or achievement, so item difficulty refers to how difficult it is for a respondent to agree with a statement, and person ability refers to the overall slant of the person's attitude, or the likelihood of the person endorsing the item, given its difficulty. Person fit is reported as person outfit and person infit, and is roughly comparable to a z-score. A mean of 0 manifests perfect fit, or response which is consistent with expectations for the respondent. Outfit is unweighted, sample-dependent, and is more sensitive to outliers than infit. Infit is weighted, independent of the sample, and less sensitive to outliers. Ideally, the distribution of item difficulty and person ability should be similar; that is, items should be provided that represent every level of agreement for the sample.

Misfit occurs when a response is not consistent with the respondent's ability, given the item difficulty. For this study, a fit statistic of greater than or equal to |2.00| was considered evidence of misfit. Positive person misfit, called underfit, indicates that the person found it difficult to respond favorably to items. Negative person misfit, called overfit, indicates the person found it too easy to respond favorably to items.

When misfit occurs, a closer examination can be made to determine the reason for the misfit, and several response sets can emerge. For example, extreme responding style is evident in the choice of only extreme responses; a slow-to-warm-up tendency is observed when responses begin erratically and then fall into a consistent pattern later on; an erratic pattern overall may signify random guessing, due to fatigue or unfamiliarity with the topic.

This study sought to identify response sets most closely associated with person fit. A common use of the Rasch model is for increasing the validity of a scale by ensuring that items fit the purpose of the scale, using both item fit and person fit statistics. However, person fit has also been shown to be an effective method for identifying response sets on a questionnaire. Its purpose in this study is to examine its potential as an indicator of the three response sets from Hui and Triandis' model (1985) -- 1) acquiescence/directional bias (A/D), 2) extreme responding style (ER), and 3) response range (RR).

Method

Sample

Subjects in this study were undergraduate and graduate college students from 11 colleges and universities in Colorado (N=597), taken from a larger study examining response set, item format, and thinking style (Swearingen, 1997). Five major areas of study (art/music, education, business, math/science, and religion) were targeted in this previous study to obtain a diverse sampling of thinking styles, with the purpose of determining if thinking style was related in some way to response set. It was concluded that there was no significant relationship between thinking style and response set for most of the response sets measured, but a possible minor association between thinking style and person fit. Additionally, Swearingen found that there are significant relationships among several of the response sets examined.

Instruments and Procedure

Subjects were administered surveys and questionnaires during class time, including two envelopes -- a white envelope containing a consent form and the Gregorc Style Delineator (Gregorc, 1984), a 4-minute, timed thinking style measure; and a yellow envelope containing 12 short attitude questionnaires covering four topics in three different item formats. The attitude measures were untimed, but were generally completed in total within 30 minutes. The topics included two controversial topics (a woman's right to an abortion, homosexual rights) and two non-controversial topics (arts education, standardized testing). The three item formats used were the semantic differential (SD), the rating scale (RS), and the magnitude estimation scale (ME). This design was an effort to control for response set due to item content, believed to be unrelated to response set, and to control for effects of item format. Attitude measures were administered in two different orders, one the reverse of the other, to control for effects of fatigue.

The SD format has been in use since the 1940s when Stagner and Osgood (1946, cited in Snider & Osgood, 1969, p. 30) conducted a study of social stereotypes. It is based on the premise that "words represent things because they produce a replica of the actual behavior toward those things, as a mediation process" (Osgood, 1952, cited in Snider & Osgood, 1969, p. 10). It consists of a series of bipolar pairs of adjectives placed on either end of a rating scale, usually with seven points in between

each pair, though some scales may have as many as 10 points. The respondent's choice of a scale-point is supposed to represent his/her feeling about the attitude object, and indicates both direction and intensity of attitude. Though items in the SD tend to produce a three-factor model, consisting of evaluative, potency and activity pairs, items for this study were selected to be evaluative pairs only, since the evaluative factor has been found highly associated with attitude (Lawson, 1989; Snider & Osgood, 1969; Tittle & Hill, 1967). The SD format is considered reliable for measuring attitudes, with studies reporting estimates of .90-.93 (Marshall & Merritt, 1986, cited in Emmerson & Neely, 1988, p. 268). A sample question from the study in the SD format looked like this:

Harmful ___ ___ ___ ___ ___ ___ ___ Beneficial

The respondent was asked to place a mark on the continuum to represent how he/she feels about standardized testing, for example.

The RS format is one of the most commonly utilized. Respondents are presented with from three to seven possible degrees of agreement for indicating how they feel about a statement. Usually, the scale-points represent choices on a continuum from strong agreement to strong disagreement. Like the SD format, it is bi-directional, indicating both direction and intensity of attitude. Tittle and Hill (1967) found greatest reliability for the RS format with 5 scale-points, though there is some controversy over the number that is most effective. A sample question on homosexual rights in the RS format was:

	Strongly Disagree						Strongly Agree
I would not hesitate to join a rally in favor of homosexual rights	1	2	3	4	5	6	7

The respondent was asked to circle the number representing his/her feeling about this statement.

In the ME format, the respondent has an opportunity to map his/her feelings on a more expansive scale. This technique was developed by Stevens (1957, cited in Schreisheim & Novelli, 1989). It may have 100 points, or 1000, or more, usually organized and labeled in ranges of 10 or more points. Though it is unidirectional, the 0 at one end actually denotes disagreement or no agreement, and the high end of the scale represents complete agreement. It is based on the assumptions that people generally are able to manipulate numbers to express ratios (e.g., if something is 100, then 200 is twice its size), and that people can perceive some kind of internal continuum which they can relate to a stimulus statement. An example of a question from the survey on arts education in the ME format was:

Art and music classes only produce restlessness in students, distracting them from academics.

0_____100_____200_____300_____400_____500_____600_____700

The respondent's mark along the continuum again represents his/her attitude about the statement.

Response sets examined in Swearingen's study (1997) were: extreme responding style (ER), response range (RR), and acquiescence/directional bias (A/D), components of Hui and Triandis' model of response sets (1985). Person fit using the Rasch model was added to augment the information derived from the Hui and Triandis model.

Scoring

ER for this study was scored by tallying the number of responses at either end of a scale for one individual. RR was determined by computing the standard deviation of a person's responses on a scale around his/her own mean for that scale. A/D was computed as the mean of an individual's responses for each questionnaire. These computations are consistent with Hui and Triandis' definitions of these sets (1985); though, they also present an alternative method for computing RR, namely subtracting the lowest response from the highest response, in addition to the standard deviation method. Response pattern (RP), represented by person fit, as stated earlier, was computed using the Rasch model on the BIGSTEPS computer program (Wright & Linacre, 1994).

Statistical Techniques

In Swearingen's study (1997), the Rasch model was applied to the data to produce person fit statistics. Then using SPSS (SPSS, Inc., 1988) correlations were computed to identify relationships among the response sets, and ANOVAs assessed effects of several variables on the incidence of response set including person fit. For the current study a closer examination was made of the Rasch output from the BIGSTEPS computer program (Wright & Linacre, 1994) for explanations of individual misfit; specifically, poorly-fitting persons, or those with underfit scores greater than 2.0. Where person fit and other response sets were found in the correlational analysis to be highly associated, verification was sought in the Rasch output. Reliability estimates of the instruments were also computed and could be compared with the person separation reliability estimates produced by the Rasch analysis.

The BIGSTEPS computer program (Wright & Linacre, 1994) eliminates "extreme" persons (those with zero or perfect scores) from the analysis. Extreme, in this sense, is different from extreme responding style, though some subjects with high ERs may be included in this group. These persons cannot be calibrated because their scores contain no information about items and ability. It cannot be known whether their

"extreme" scores are a result of response set or whether items were too hard or too easy for them, or whether their responses truly represent agreement or disagreement. This meant that for the analysis of some scales, there were many fewer subjects than the 569 which the final sample provided, after persons with invalid surveys were dropped.

Results

The 569 subjects for this study included 43.9% males and 55.5% females. The average age of the sample was 28, with 70% of the sample under age 30, though ages ranged from 17 to 61. Ethnicity categories were unbalanced, with 78.3% classifying themselves as Anglo-American, 7.2% as International students, and 4.8% as Hispanic-Americans. Other ethnicity groups were in even smaller number.

Reliability estimates from the SPSS program (SPSS, Inc., 1988, 1994) are shown in Table 1. The SD scale maintained highest reliability across formats and content areas, consistently above .90. This is commensurate with the studies of Marshall and Merritt (1986, cited in Emmerson & Neely, 1988) that found high reliability estimates for SD scales. The ME scale was found least reliable overall, and non-controversial content areas were less reliable than controversial ones for the RS and ME scales. Unfamiliarity with the ME format and difficulties in interpretation of subject's responses may be responsible. The locations for some subjects' responses along the continuum were unclear. It may be also that with different topics, different results may be seen. Further study could examine the role of fit statistics in explaining reasons for reliability differences among formats and content areas.

Table 2 displays category response frequencies for each item on each of the 12 scales. A glance can inform that with some of the scales responses to questions were highly skewed; whereas, with others there was a more normal distribution of response. It would be expected from these distributions that ER, A/D, and RR may be detected.

Response set means across the 12 scales exhibited different patterns for each of the response sets (see Figures 1 through 4). The A/D set followed similar curves for all three formats, with the highest means occurring with the arts education scales in most formats. The RR set varied by format, with the widest divergence in response set means on the arts education and homosexual rights scales. ER exhibited peaks and valleys corresponding to the other response sets across the RS and ME formats, but the arts education scale produced a wide divergence of ER across formats. Person fit means deviated only slightly from perfect fit, but the widest range of misfit occurred with the SD scale on standardized testing.

Person infit means for the 12 attitude scales ranged from -.22 to -.81, indicative of only slight deviation from perfect fit overall. However, standard deviations revealed wide ranges of individual infit means (s.d. range, 1.01 to 1.62).

Person infit and person separation reliability information are displayed in Table 3. Reliability estimates from this analysis are based on non-extreme persons only, so may be seen as more informative or more useful than traditional reliability estimates that include perfect and zero scorers. Since statements about measures for these latter persons are considered imprecise, their data may be said to contaminate traditional reliability estimates. The Rasch reliability estimate is similar to a KR20, and the SPSS estimate is a Cronbach's alpha.

The analysis of association among response sets (see Table 4) indicated low-moderate to moderate, positive correlations between person infit and ER for all four topics on the SD and RS scales ($\underline{r}$ = .34-.63). Moderate to substantial, positive correlations were found between person infit and the RR response set for all four topics on the SD scale ($\underline{r}$ = .50-.88), for all but the homosexual rights scale in RS format ($\underline{r}$ = .46-.73), and for the standardized testing scale in the ME format ($\underline{r}$ = .65). A/D was not significantly associated with person infit.

Results of the correlational analysis also revealed very high associations between infit and outfit ($\underline{r}$ = .93-1.00), indicating redundancy. The infit statistic was chosen, then, as a measure of RP since it is relatively unaffected by outliers. The infit statistic gives the added information that the person responded unexpectedly to items near his/her ability level (Linacre & Wright, 1997). It is this kind of response that would signal incidence of response set.

Figures 5 through 8 give maps of persons and items for four of the attitude scales, providing a clear visual representation of the degree of alignment of items with persons, based on item difficulty and person ability. The first column shows the distribution of persons by ability along the vertical logit scale. The second column indicates the placement of the lowest item responses along the same scale; the third column locates the mid-range item responses; and the last column places high item responses. When item responses are above or below the person distribution, they are either too hard or too easy for the sample. When persons have no items matching their location on the logit scale, then no items exist on the scale to measure their attitudes at all levels. This weakens the usefulness of the scale for those people, and items are considered to be poorly designed for the sample. On the SD scale on arts education, for example, too many of the sample are above the scale of items, so the scale cannot successfully measure the attitudes for those persons. For the SD on homosexual rights, again there is a large portion of the sample above the items, but middle item responses are better centered within the middle ability groups, so those groups are measured fairly well. For the RS on abortion rights, both low and high item responses lack persons to measure. The map for the ME scale on standardized testing is closer to what is expected. The sample is fairly normal, though it has both a long positive and a long negative tail. Middle item responses are centered well with the sample, and high and low item responses measure persons in the tails of the distribution, but there are insufficient numbers of persons in the tails to be measured at high and low attitude.

A closer look at individual output on the Rasch analysis permitted an observation of responses of misfitting persons, and suggested specific reasons for misfit. Figure 9 gives examples from the output of some of the most misfitting persons' responses to items on the SD on arts education, with items arranged in ascending order by difficulty. It can be seen that a majority of the misfitting persons had poor fit because of extreme responses or because of wide response range. Their responses were not consistent with their ability and the item's difficulty. For example, person #422 (infit=5.2) responded with all 7's, indicating extreme positive response, except to one item. According to this person's ability (1.61 logits), s/he should find it easy to agree, but the response to the second item is an extreme negative one. Person #447, with an ability of .00 (infit=2.5) is expected to respond with a 50% chance of agreeing or disagreeing. But this person responded with fairly strong agreement and disagreement to the items.

These observations verify what the high correlations between infit and RR and infit and ER suggested. Person fit can be useful in detecting ER and RR response sets. A/D is not as easily detected by the Rasch analysis, because a person with all agree or all disagree responses is eliminated from the analysis.

Because of the short length of the individual questionnaires in this study, fatigue was not evident or easily observed; though it may be observed due to repeating topics in different formats. Random guessing may be suggested by the patterns of persons #447, #495, and #367, whose responses seem to cover all item-response ranges.

Discussion

Limitations

The small number of items per scale in this study limited the ability to detect a wider variety of sets than might be possible with lengthier scales. It was also difficult to equate items across formats, since the semantic differential involves word-pairs, and the other scales involve statements. A better comparison could be made of response set across formats if the formats used had items that were parallel.

Because the sample was comprised of college students, the sample was perhaps more motivated than some persons would be in responding. However, because they came from intact classes, a few may have felt trapped and unable to decline participation in front of their peers, even though participation was voluntary. This can increase the likelihood of extreme misfitting responses. The exclusion of extreme persons made it difficult to detect some response sets, such as A/D and extreme responding style for such persons.

Conclusions

The moderate-to-substantial correlations between infit and ER and between infit and RR found on the SD and RS scales are not seen for the ME scale, suggesting fit

statistics may be useful in determining response set on the SD and RS scales for all but the A/D set, and perhaps not as consistently useful with the ME scale. It is especially interesting to note that associations of response sets with infit averaged higher than associations among any other response set pairs, a strong suggestion that person fit statistics deserve more attention in response set research.

Because of the high associations observed, the measurement of person fit through use of the Rasch model is an effective method for detecting response set. In particular, it detects RR and ER very quickly, and perhaps random guessing, even on a scale with few items. On a larger scale, it is expected that random guessing would be more apparent, as would slow-to-warm-up tendencies, and fatigue. A/D is not as easily seen from the Rasch analysis. So many models have been devised to identify response set, but it may be that the Rasch model will be seen as a device to detect a wider variety of sets in one analysis, without the need for separate computations for each one. It is noteworthy that the substantial correlations found in this study between person fit and other response sets indicate also that person fit detects response set irrespective of item format, since these correlations were found in most formats.

The SD on arts education was found to have a poor set of items for the sample measured (See Figure 5). A look at the frequency distributions of A/D, RR, and ER (Figures 1, 2, and 3) indicates wide departures for the SD scale on these response sets. This can be seen also for the SD scale on homosexual rights (Figures 2 and 6). In addition to providing another means for detection of response inconsistencies, analysis of person fit adds legitimacy, then, to response sets detected by other means.

.Response set is an ever-present phenomenon threatening the accuracy of information we derive from measurement. An awareness of this and the ability to detect its many forms is a high priority for communicators of test and survey results. The Rasch model provides information for accomplishing this in the form of person fit scores.

REFERENCES

Alwin, D. F., & Krosnick, J. A. (1991). The reliability of survey attitude measurement. Sociological Methods and Research, 20(1), 139-181.

Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: Black-white differences in response styles. Public Opinion Quarterly, 48, 491-509.

Beardon, W. O., & Rose, R. L. (19901). Attention to social comparison information: An individual difference factor affecting consumer conformity. Journal of Consumer Research, 16, 461-471.

Cantril, H. (1946). The intensity of an attitude. Journal of Abnormal and Social Psychology, 41, 129-135.

Cronbach, L. J. (1946). Response tests and test validity. Educational and Psychological Measurement, 6, 474-494.

Cronbach, L. J. (1950). Further evidence on response sets and test design. Educational and Psychological Measurement, 10, 3-31.

Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. Journal of Applied Psychology, 37(2), 90-93.

Emmerson, G. J., & Neely, M. A. (1988). Two adaptable, valid, and reliable data-collection measures: Goal attainment scaling and the semantic differential. The Counseling Psychologist, 16(2), 261-271.

Gregorc, A. F. (1984). Gregorc Style Delineator: Development, technical and administration manual. Columbia, CT: Gregorc Associates, Inc.

Hamilton, D. L. (1968). Personality attributes associated with extreme response style. Psychological Bulletin, 69(3), 192-203.

Hui, C. H., & Triandis, H. C. (1985). The instability of response sets. Public Opinion Quarterly, 49, 253-260.

Hui, C. H., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. Journal of Cross-Cultural Psychology, 20(3), 296-309.

Lawson, E. D. (1989). Sex-related values and attitudes of college students: A sexism scale vs the semantic differential. Psychological Reports, 64, 463-476.

Linacre, J. M., & Wright, B. D. (1997). A user's guide to BIGSTEPS. Chicago, IL: Mesa Press.

McClendon, M. J. (1991). Acquiescence and recency response-order effects in interview surveys. Sociological Methods and Research, 20(1), 60-103.

Meisels, M., and Ford, L. H., Jr. (1990). Social desirability response set and semantic differential evaluative judgments. The Journal of Social Psychology, 78, 45-54.

Rorer, L. G. (1965). The great response style myth. Psychological Bulletin, 63(3), 129-156.

Schreisheim, C. A., & Novelli, L. (1989). A comparative test of the interval-scale properties of magnitude estimation and case III scaling and recommendations for equal-interval frequency response anchors. Educational and Psychological Measurement, 49(1), 59-74.

Smith, R. M. (1986). Person fit in the Rasch model. Educational and Psychological Measurement, 46, 359-372.

Snellbecker, G. E. (1993, April). Individual differences in response consistency-inconsistency -- Measurement artifacts or psychological phenomena? Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

SPSS, Inc. (1988). SPSS-X user's guide, 3rd edition. Chicago, IL.

Snider, J. G., & Osgood, C. E. (Eds) (1969). Semantic differential technique: A sourcebook. 3-10, 21-82, 161-168, 467-473, 625-636.

Swearingen, D. L. (1997). Response sets, item format, and thinking style: Implications for questionnaire design. Dissertation Abstracts International, 98, 04094.

Tittle, C. R., & Hill, R. J. (1967). Attitude measurement and prediction of behavior: An evaluation of conditions and measurement techniques. Sociometry, 30, 199-213.

White, B. J., & Harvey, O..J. (1965). Effects of personality and own stand on judgment and production of statements about a central issue. Journal of Experimental Social Psychology, 1, 334-347.

Wilcox, C., Sigelman, L., & Cook, E. (1989). Some like it hot: Individual differences in response to group feeling thermometers. Public Opinion Quarterly, 53, 246-257.

Wright, B., & Linacre, R. (1994). A user's guide to BIGSTEPS: Rasch model computer program. Chicago, IL: Mesa Press.

Wright, B. D., & Stone, M. H. (1988, March). Reliability in Rasch measurement. Research Memorandum No. 53. Chicago, IL: MESA Psychometric Laboratory.

Table 1

Reliability Estimates for the 12 Attitude Scales (N=548)

Scale	SD	RS	ME	Topic Mean Alpha
Woman's Right to an Abortion	.94	.82	.92	.89
Arts Education	.95	.73	.67	.78
Homosexual Rights	.96	.82	.78	.85
Standardized Testing	.93	.73	.59	.75
Format Mean Alpha	.95	.78	.74	.82

Note: SD - semantic differential
RS - rating scale
ME - magnitude estimation

Table 2

Response Frequencies and Item Difficulty for Each of the 12 Attitude Scales

Scales/Items	Response Categories 1	2	3	4	5	6	7	Estimated Item Difficulty
SD on Abortion Rights								
Item 1	46	46	21	37	46	96	148	−.55
2	68	57	45	94	56	89	31	.37
3	69	69	45	110	45	66	36	.43
4	55	53	39	116	52	89	36	.22
5	34	34	24	111	38	93	106	−.47
SD on Arts Education								
Item 1	9	15	6	36	53	132	93	−.21
2	10	13	12	28	43	138	100	−.23
3	11	16	13	35	60	127	81	.00
4	4	14	17	63	46	136	64	−.19
5	20	20	18	73	49	119	45	.64
SD on Homosexual Rights								
Item 1	42	53	32	40	51	95	88	−.22
2	42	57	43	79	55	91	34	.30
3	42	56	55	87	49	80	32	.40
4	35	45	43	81	56	86	55	.00
5	24	39	34	72	46	87	99	−.49
SD on Standardized Testing								
Item 1	2	30	64	66	126	98	21	−.34
2	2	36	72	80	132	98	21	−.19
3	3	55	99	94	139	57	17	.41
4	3	28	68	76	155	94	20	−.11
5	3	61	81	96	120	66	27	.24
RS on Abortion Rights								
Item 1	110	44	22	14	38	101	184	.15
2	53	43	53	86	47	99	132	.08
3	24	23	28	57	52	104	225	−.39
4	79	50	36	27	39	68	214	.02
5	46	44	47	78	96	108	93	.13
RS on Arts Education								
Item 1	6	9	16	9	53	103	349	−.71
2	36	22	28	250	92	89	28	.87
3	14	12	11	25	92	181	209	−.30
4	18	14	20	247	77	113	56	.47
5	8	11	30	67	67	157	204	−.33
RS on Homosexual Rights								
Item 1	32	29	40	66	48	101	194	−.46
2	85	53	33	25	42	102	171	−.11
3	13	16	29	26	32	112	283	−.93
4	120	51	45	70	43	85	95	.30
5	208	55	40	97	47	42	22	1.19

table continues

Table 2 - continued

Scales/Items	Response Categories 1	2	3	4	5	6	7	Estimated Item Difficulty
RS on Standardized Testing								
Item 1	132	167	117	42	36	38	18	.31
2	44	88	93	73	131	91	30	−.32
3	59	85	79	170	77	61	19	−,10
4	61	107	88	115	103	59	16	−.04
5	104	118	103	130	35	39	21	.14
ME on Abortion Rights								
Item 1	89	65	38	37	33	62	131	.33
2	77	49	19	38	41	85	146	.11
3	39	61	31	40	42	86	156	−.15
4	61	46	29	35	55	85	144	.01
5	32	31	17	70	66	110	128	−.30
ME on Arts Education								
Item 1	25	32	38	143	120	119	71	.25
2	3	11	14	26	61	144	290	−.79
3	9	14	10	16	33	100	367	−.67
4	46	41	42	119	116	130	55	.46
5	43	32	42	239	81	86	24	.75
ME on Homosexual Rights								
Item 1	77	72	68	73	46	59	78	.44
2	25	39	44	61	55	108	141	−.08
3	10	20	20	26	42	98	257	−.53
4	47	39	44	54	50	81	158	.04
5	74	57	34	54	24	49	181	.14
ME on Standardized Testing								
Item 1	52	136	95	138	68	52	21	.32
2	36	84	75	148	86	84	49	−.02
3	84	92	97	118	59	62	48	.22
4	20	36	59	117	125	110	96	−.38
5	25	57	66	140	116	113	45	−.13

Note: SD - semantic differential
RS - rating scale
ME - magnitude estimation

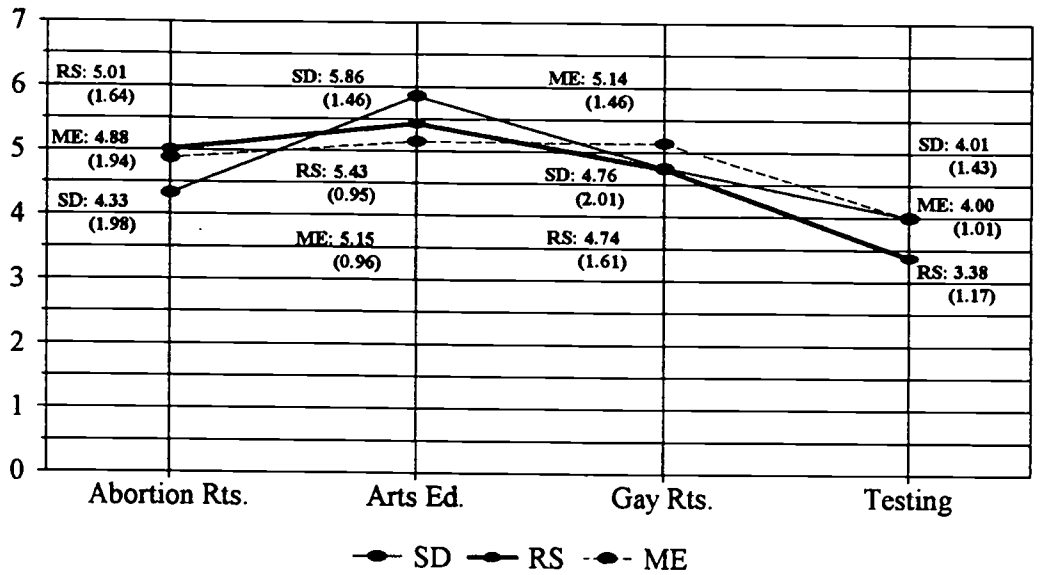

Figure 1. Acquiescence/Directional Bias Means and Standard Deviations ( ) for the 12 Attitude Scales

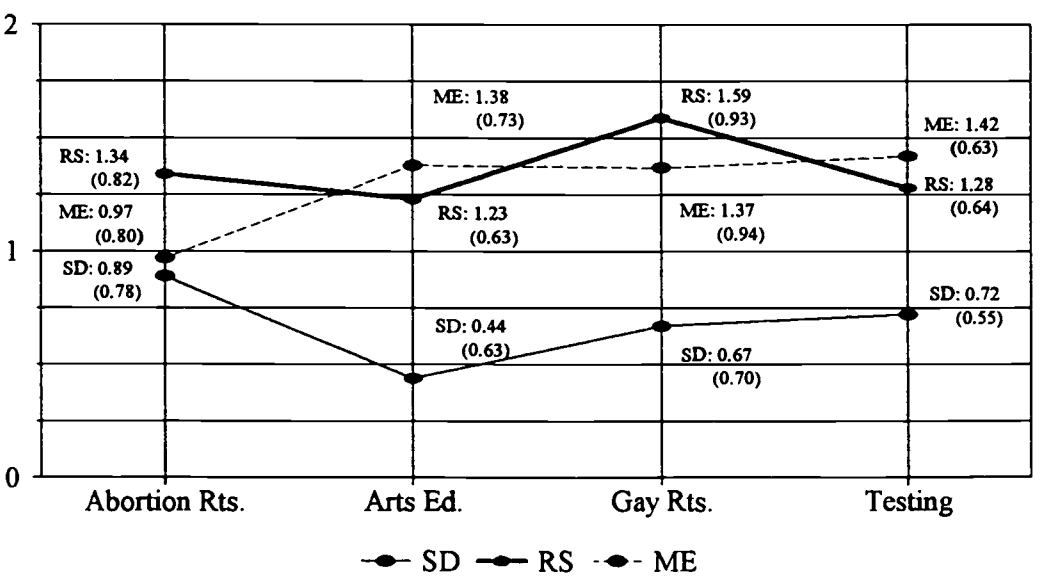

Figure 2. Response Range Means and Standard Deviations ( ) for the 12 Attitude Scales

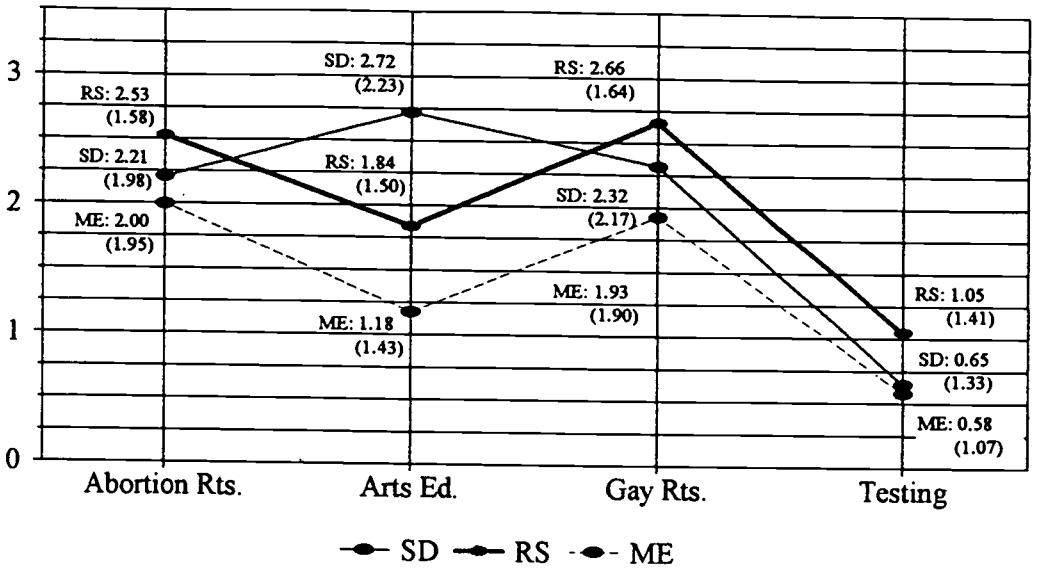

Figure 3. Extreme Responding Style Means and Standard Deviations ( ) for the 12 Attitude Scales

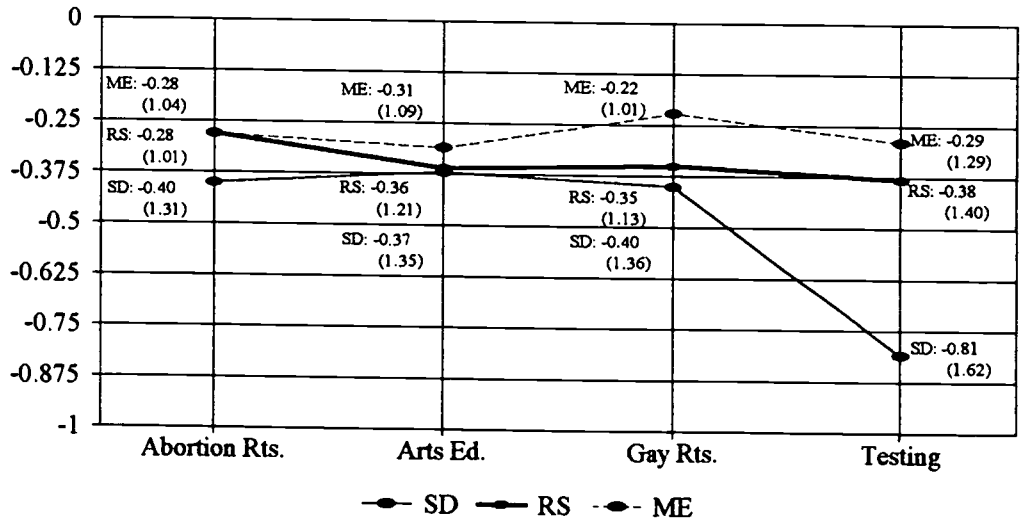

Figure 4. Standardized Person Infit Means and Standard Deviations ( ) for the 12 Attitude Scales

Table 3

Person Information on Poorly Fitting Persons and Separation Reliability by Attitude Scale

Scale	Person Statistics							
	Analyzed N	Item N	Person Ability Mean	Person Ability s.d	Person Infit Mean	Person Infit s.d.	Person Infit # >2.0	Person Separation Reliability
SD--Abortion Rights	440	5	.19	1.36	–.5	1.5	30	.87
Arts Education	344	5	1.47	1.68	–.6	1.7	36	.85
Homosexual Rights	401	5	.39	1.67	–.6	1.6	25	.89
Standardized Testing	554	5	1.08	2.14	–.8	1.6	31	.92
RS--Abortion Rights	513	5	.43	.79	–.3	1.1	10	.68
Arts Education	545	5	.94	1.07	–.4	1.2	29	.74
Homosexual Rights	511	5	.35	1.06	–.4	1.2	27	.80
Standardized Testing	550	5	–.39	.79	–.4	1.4	37	.74
ME--Abortion Rights	455	5	.44	1.12	–.3	1.2	14	.79
Arts Education	549	5	.66	.89	–.3	1.2	20	.71
Homosexual Rights	473	5	.45	.67	–.3	1.2	23	.64
Standardized testing	563	5	.05	.61	–.4	1.5	39	.66

Note: SD -- semantic differential
RS -- rating scale
ME -- magnitude estimation

Table 4

Correlations among Response Set Variables

Format	SD				RS				ME			
Content →	Abor. Rts.	Arts Ed.	Gay Rts.	Test.	Abor. Rts.	Arts Ed.	Gay Rts.	Test.	Abor Rts	Art Ed.	Gay Rts.	Test.
Response Set Pairs ↓												
ER,RR	-.24	-.23	-.30	-	-.17	-	-	.14	-.20	.21	-.17	.38
ER,A/D	-	.49	.26	-.11*	.33	.57	.26	-.44	.17	.32	.43	-
ER,Infit	.58	.59	.56	.34	.63	.44	.43	.58	.21	-	.19	.22
ER,Outfit	.57	.59	.55	.34	.57	.44	.26	.57	.21	-	.17	.22
RR,A/D	-	-.32	-.27	-	-.44	-.48	-.47	.23	-.27	-.51	-.57	-
RR,Infit	.50	.57	.54	.88	.46	.52	.22	.73	.24	.19	.20	.65
RR,Outfit	.49	.56	.53	.87	.47	.51	.11*	.71	.24	.14	.17	.64
A/D,Infit	-	.21	-	-	-	-	-	-.09	.14	-	-	-
A/D,Outfit	-	.20	-	-	-	-	-	-.11*	.13*	-	-	-
Infit,Outfit	.99	1.00	1.00	1.00	.97	.97	.93	.99	.99	.94	.97	1.00

Note: SD - semantic differential RS - rating scale ME - magnitude estimation
ER - extreme responding style RR - response range
A/D - acquiescence/directional bias Infit - standardized person infit
Outfit - standardized person outfit * $p \leq .01$ "-" $p > .05$
All correlations have a significance level of ≤.001, unless otherwise noted.

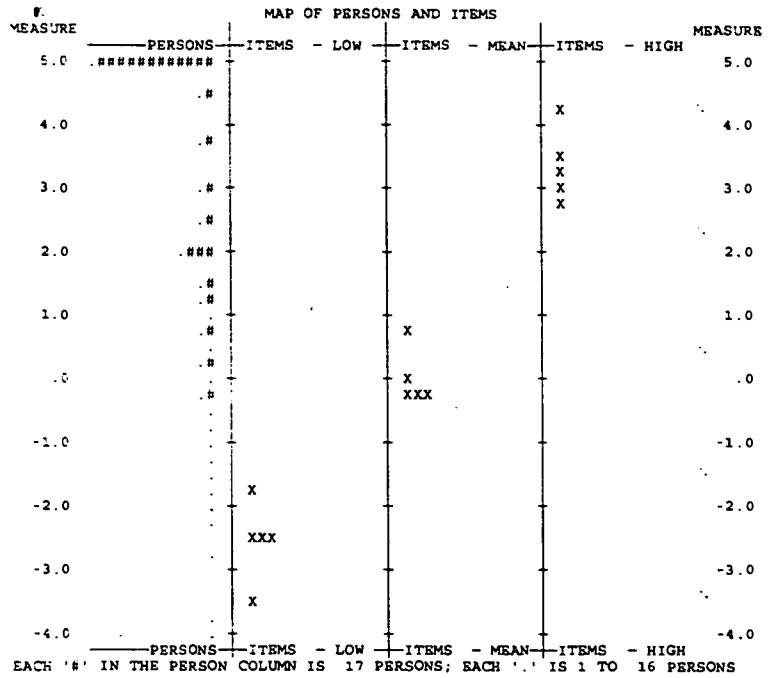

Figure 5. Map of Persons and Items for the Semantic Differential on Arts Education

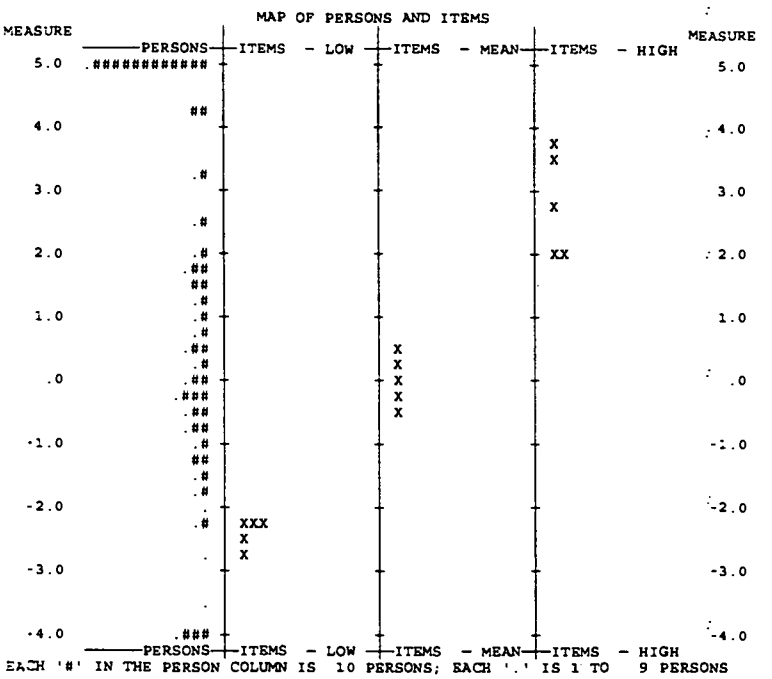

Figure 6. Map of Persons and Items for the Semantic Differential on Homosexual Rights

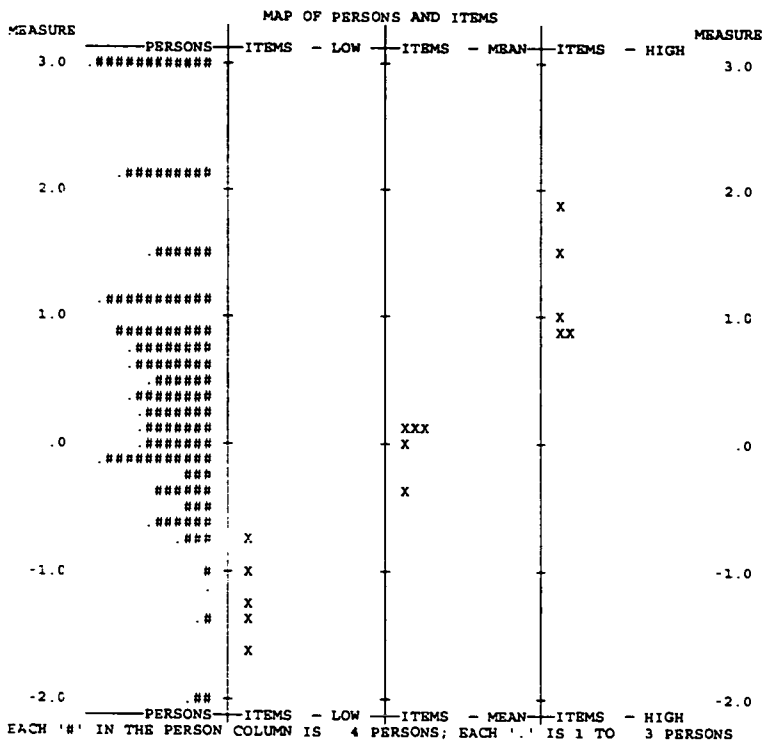

Figure 7. Map and Persons and Items for the Rating Scale on Abortion Rights

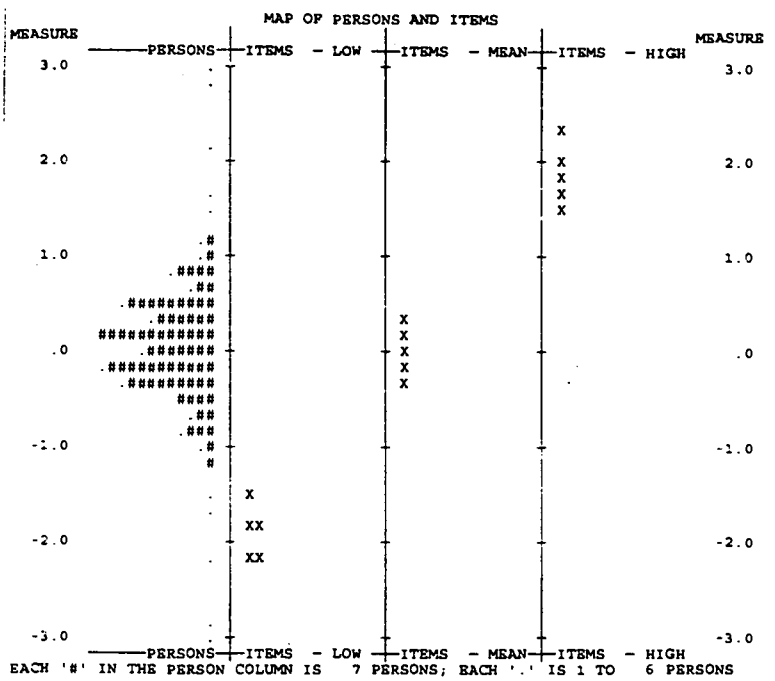

Figure 8. Map of Persons and Items for the Magnitude Estimation Scale on Standardized Testing

NUMBER – NAME — POSITION	MEASURE	INFIT	(ZSTD)	OUTFIT
400 422	1.61	5.2	A	5.7
RESPONSE: 1: 7 1 7 7 7				
RESIDUAL: -7				
485 509	1.18	4.4	B	4.0
RESPONSE: 1: 7 7 M 1 7				
RESIDUAL: -4 2				
470 493	-.70	4.0	C	3.9
RESPONSE: 1: 7 7 1 1 1				
RESIDUAL: 2 2 -2				
450 473	.17	3.8	D	3.8
RESPONSE: 1: 7 7 1 7 1				
RESIDUAL: 2 -3 2-2				
549 575	.17	3.8	E	3.8
RESPONSE: 1: 7 7 1 7 1				
RESIDUAL: 2 -3 2-2				
187 197	1.61	4.3	F	3.5
RESPONSE: 1: 7 7 7 7 1				
RESIDUAL: -4				
314 329	1.61	4.3	G	3.5
RESPONSE: 1: 7 7 7 7 1				
RESIDUAL: -4				
393 415	1.61	4.3	H	3.5
RESPONSE: 1: 7 7 7 7 1				
RESIDUAL: -4				
526 551	1.61	4.3	I	3.5
RESPONSE: 1: 7 7 7 7 1				
RESIDUAL: -4				
457 480	2.44	3.3	J	3.5
RESPONSE: 1: 7 7 7 3 7				
RESIDUAL: -4				
388 409	-1.09	2.2	V	2.6
RESPONSE: 1: 2 2 2 2 6				
RESIDUAL: 3				
425 447	.00	2.5	W	2.5
RESPONSE: 1: 7 6 2 6 1				
RESIDUAL: 2 -2 -2				
472 495	.17	2.1	X	2.2
RESPONSE: 1: 2 6 6 3 6				
RESIDUAL: -2				
349 367	.77	2.0	Y	2.1
RESPONSE: 1: 6 3 7 4 6				
RESIDUAL: -3				
6 6	2.99	2.3	Z	2.1
RESPONSE: 1: 7 7 7 7 4				
RESIDUAL: -3				

Figure 9. Examples of Misfitting Persons' Responses to Items on the Arts Education Semantic Differential Scale

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM028994

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: PERSON FIT AND ITS RELATIONSHIP WITH OTHER MEASURES OF RESPONSE SET

Author(s): DOROTHY L. SWEARINGEN, PH.D.

Corporate Source: (Presented at Annual Meeting of American Educational Research Association)

Publication Date: APRIL, 1998

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☒

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Dorothy L. Swearingen, Ph.D.	Printed Name/Position/Title: DOROTHY L. SWEARINGEN, PH.D.
Organization/Address: 1440 S. GARFIELD ST. DENVER CO 80210	Telephone: (303) 756-2982 FAX: —
	E-Mail Address: dswearin@ix.netcom.com Date: 5/18/98

(over)